

An Update on Development of the Vireo 4.x ETD Submission and Management System

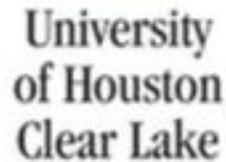
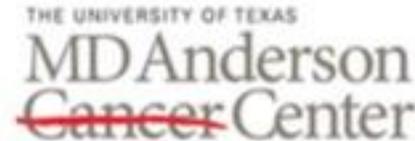
A seminar on the project background, development progress, community involvement, and release scheduling presented at the Texas Conference on Digital Libraries 2017, Austin Texas

James Creel, Jeremy Huff, Jason Savell, William Welling, Ryan Laddusaw, Douglas Hahn, Michael Bolton, Ryan Steans, and Stephanie Larrison

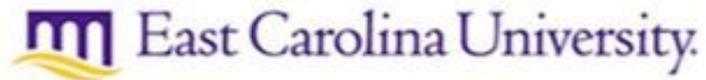
Seminar Outline

- A Brief Overview of Vireo
- Project Motivations
- Notes from Last Year
- Workflow Management Features
- Organizational Hierarchy Features
- Administrative Features
- Release Scheduling
- How to Get Involved

Vireo @ TDL



Other Vireo Deployments



Management and Oversight (1/3) – The VUG

VIREO USERS GROUP



- Votes on new features
- Prioritizes features and casts as user stories
- Provides a Product Owner for Agile development with Vireo developers

Management and Oversight (2/3) – The TDL ETD Metadata Working Group

- Based on extensive research of ETD metadata usage across the state, in 2015 this group published a set of recommendations to be implemented in new Vireo versions.
- Texas Digital Library Descriptive Metadata Guidelines for Electronic Theses and Dissertations, Version 2.0:
<http://hdl.handle.net/2249.1/68437>

Management and Oversight (3/3) – TDL, TAMU, and VUG doing Agile development

- Stephanie Larrison, co-chair of the VUG is serving as product owner for Vireo 4.
- The product owner chooses the work to be done on a given development sprint and approves completed work for inclusion in the code base
- TDL, TAMU, and Texas Tech have all contributed developer effort toward the 4.0.x code base with development sprints being run by TAMU.

A Brief History of Vireo

- 2009: **Vireo 1.0**. A heavily customized DSpace workflow
- 2012: **Vireo 1.8**. A complete rewrite in the Play! framework
- 2013: **Vireo 2.0**. New fields, email workflow rules, batch operations, exports (including ProQuest)
- 2015: **Vireo 3.0**. ORCID support, new fields, email and workflow tweaks
- Meanwhile, as Vireo grows more established and widely used, the VUG starts collecting user stories that would stretch the fixed field set of Vireo beyond the limit...
- Summer 2015: Vireo 4.0 is conceived as a Spring Boot and Angular.js application with fully customizable workflows and configurable controlled vocabularies

Project Motivations (1/5)

Controlled Vocabularies for Advisor Names

- Graduate office employees have to spend time normalizing advisors' names that are input as free text
- Authorities are readily available to control these: lots of agencies on campus, including the registrar and the payroll office know all the faculty's official names and email address

Project Motivations (2/5)

Controlled Vocabularies for Subject Keywords

- Subject keywords entered by students as free text are of poor quality despite the students' mastery of their subject
- Library scientists have developed ample thesauri and subject heading lists for all established disciplines

Project Motivations (3/5)

Controlled Vocabularies for Linked Open Data

- Advisor name disambiguation can be used in conjunction with institutional URIs (such as VIVO profiles) or ORCID.
- Most subject authorities (LOC authorities, VIAF, NALT, Geonames) provide URIs for their constituent terms.
- In both cases, referents of URIs can participate in the ecosystem of linked open data

Project Motivations (4/5)

Workflows for Different Programs

- Different degree programs may require different workflows
 - ProQuest metadata
 - Embargo options
 - Available disciplines
 - Advisory committee structure
 - Document structure

Project Motivations (5/5)

Workflows for Different Disciplines

- Different disciplines have different subject matter that can be captured in descriptive metadata
- Geology, Geography, History: GeoNames, Pleiades
- Agriculture: NALT
- Computer Science: ACM
- Medicine: MESH
- Art, Architecture: Getty's AAT

Notes from Last Year

- Release deadline at last year was for August 2016
- Now we're planning a beta release for August 2017
- What happened?
 - Understaffing
 - Code reusability from Vireo 3 was almost nil.
 - New features much bigger than planned
 - Controlled vocabulary interface accommodating external and internal sources equally – three iterations
 - New workflow management – untold iterations
 - Dynamically generated query for all the custom metadata – two iterations

Workflow Management in Vireo 4

- Administrators can create any number of Organizations.
- Each Organization can have it's own Workflow.
- By default, a new Organization's Workflow looks like the Vireo 3 workflow.
- Workflows consist of any number of Workflow Steps.
- Workflow Steps consist of any number of Field Profiles and Notes
- Field Profiles characterize the mode of input for a given metadatum on a submission.
- Notes are helpful text that appears alongside the inputs on a step.

Field Profiles

- Basic settings:
 - Label
 - Repeatable
 - Required
 - Disabled
- Controlled Vocabulary
- Advanced Configuration
 - Types: text, email, contact, password, checkbox, datetime, file, radio, telephone, URL, textarea, select (i.e. dropdown), ORCID, date
 - Flagged for prominent display
 - Logged
 - Overrideable
 - Metadata Key/Document Type
 - Help text
 - Shibboleth attribute mapping

Submissions are made according to their Organization's Workflow

- A student starts a Submission by selecting the Organization to which they are submitting.
- When the new Submission is created, it takes a snapshot of the Organization's workflow at the time of creation.
- If an Organization changes its workflow, perhaps in a later semester or under new management, the Submission keeps the original Workflow under which it was created

Organizational Management in Vireo 4

Approaches to Organizing Workflows

- Organizations can be organized in a parent/child hierarchy according to any preference whatsoever.
- None – the hierarchy is totally optional!
- Degree/Program – e.g. graduate/undergraduate, masters/doctoral
- Organizational Hierarchy – mirror the org chart or the course catalog
- Hybrid approaches are entirely possible and likely to occur.

No Organizational Hierarchy

- The workflow for all Submissions will be the same, and the Vireo 4 instance will behave like a super-customizable Vireo 3 instance.
- This structure is our likely target for automatic upgrade/migration from a Vireo 3 instance to a Vireo 4 instance

Organization Workflows based on Program/Degree

- TAMU currently runs two Vireo instances, one for the Office of Graduate Studies, and one for the Undergraduate Research Scholars program.
- It might make sense to have one workflow for Masters degrees, and another for Doctoral degrees.

Organization Hierarchy based on Institutional Structure

- E.g. an Organization for each college
 - Each college contains an Organization for each department
 - Each department contains an Organization for each degree it grants
 - Students would pick their college, department, and degree to start the Submission

Hybrid Organizational Structures

- Likely to appear in the real world as institutions get familiar with the power of the system and want to incrementally leverage more and more flexibility
- No hierarchy or a simple one will probably be the way institutions pilot Vireo 4, but particular stakeholders will clamor for customizations once the product is in use

Inheritance of Workflow Components in the Organizational Hierarchy

- Organizations are created with the exact same workflow as their parent, and their workflows will change when the parent changes
- Organizations may override (i.e. customize) any workflow component they inherited from their parent *provided* the parent allows it
- Organizations that choose to make overrides to components no longer get the changes their parent makes *unless* the parent says that the component is *non-overrideable*

New Administrative Features

- Controlled vocabulary management
- Shibboleth attribute mappings
- Metadata key management
- Email templating with dynamic fields
- Submission list with dynamic fields

Controlled Vocabulary Management

- Create controlled vocabularies with any name and number of terms
- System configuration can additionally allow for controlled vocabularies fabricated from system Java entities
- Populate/modify the terms via CSV upload
- Future work: Full CRUD UI for the terms in any vocabulary – this will recapitulate the look and feel of the configurable settings subsumed by controlled vocabulary, e.g. degree, program, etc.

Shibboleth Attribute Mappings

- Allows the association of Weaver framework-level user associated keys with human readable labels for use in customizing fields
- Values are fed through by Weaver framework, and could actually have come from Shibboleth header or from user registration over email

Managing Metadata Keys

- Metadata keys, aka predicates, are implicated in the configuration of Field Profiles and thence in exports.
- New ones can be created during Field Profile customization
- A dedicated interface offers CRUD for predicates

Email Templates

- Templating done with metadata keys
- Rules cascade to child organizations

Submission List with Custom Filters

- Recapitulates functionality of Vireo 3, except **with a dynamic, customizable set of columns and filters**
- List columns can be customized
- Filters can be customized and saved
- Flagged metadata keys are displayed on the sidebar for quick access

Demonstration of the Submission Workflow

- Administrator sets up system for submissions
- Student conducts the initial submission
- Advisor approves submission
- Thesis office returns for corrections
- Student makes corrections
- Thesis office published to Vireo

Release Scheduling for Vireo 4

- *Alpha Release*: All features have received attention, but some remain incomplete. There are known and unknown bugs.
- *Beta Release*: Feature frozen. Ready for realistic testing to hunt down bugs. A viable app for the intrepid and adventurous.
- I consider the Vireo 4.0.x branch on Github right now an advanced *Alpha Release*. The *Beta* is planned for official announcement at the end of summer 2017.

Shameless Plug for Community Involvement

- We can't do it without you!
- Texas Tech, Baylor, and Texas State have all stepped up to the plate with generous contributions
- Development time is welcome, but there are a variety of other **equally important** ways to contribute
 - Writing exporters (not heavy duty coding, just templating)
 - Documentation
 - Testing, testing, and more testing
 - Getting buy-in for pilots from local thesis offices and students, who will do the real-world testing and find the sneakiest bugs